Automated Workflows For Structure Determination By Single-Particle Cryo-EM At Sub-nanometer Resolutions

Alberto Bartesaghi[1], Jason Pierson[2], Michael Alink[2], David Schauder[1], Soojay Banerjee[1], Prashant Rao[1], Mario J. Borgnia[1], Jacqueline Milne[1], Martin Kessel[1], Lesley Earl[1], and Sriram Subramaniam[1]

[1]Laboratory of Cell Biology, Center for Cancer Research, National Cancer Institute, NIH, Bethesda, MD USA [2]FEI, Eindhoven, Netherlands

Abstract

The combination of hardware advances in transmission electron microscopes together with automated data collection and streamlined image processing routines have transformed the pace at which samples can be converted into 3D structures by single-particle cryo-EM. In this work, we present automated workflows for data collection and processing of large single-particle datasets. Efficient micrograph screening routines together with automatic particle picking strategies enable processing of entire datasets in high-throughput mode in a routine manner. We use this platform as a framework to perform comparative studies of the effects of different types of detectors, acceleration voltages and different dose fractionation schemes on the resolution of reconstructions. We applied these workflows to multiple datasets of GroEL and β-galactosidase imaged under a variety of conditions showing that reconstructions at sub-nanometer resolution can be consistently achieved using this framework.

Need for High-Throughput Processing in Single-Particle Cryo-EM

- **Speedups in Data Collection**
  - Hardware advances in transmission electron microscopes
  - Automated data collection software
- **Advances in Data Processing**
  - Streamlined image processing routines
  - Availability of Computational Power

The Single-Particle Workflow

Automated Data Collection

↓

CTF Estimation
Automated Particle Picking
Micrograph Screening

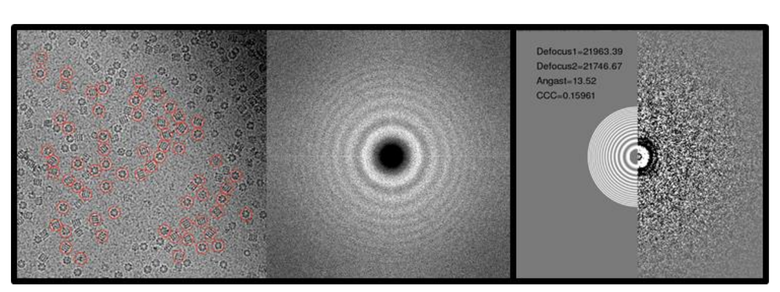
↓

3D Orientation Assignment

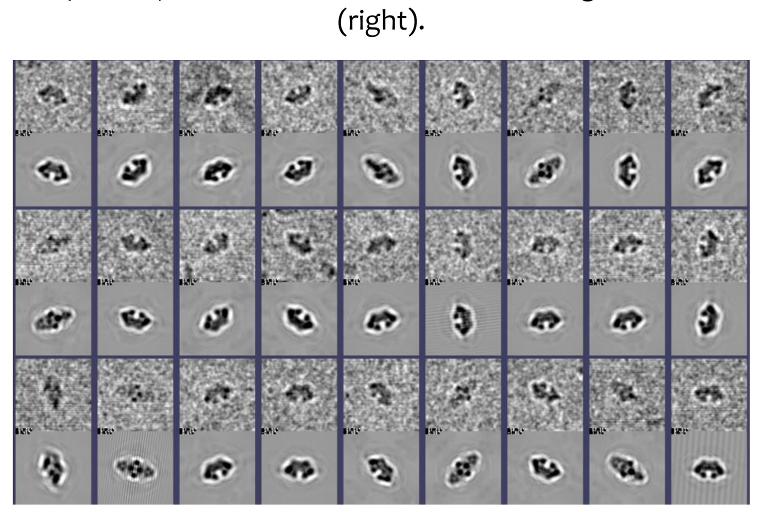
↓

Structure

EPU: Automated Single Particle Acquisition Software



Automatically picked particles (left), 2D Power Spectrum (middle) and result of CTF estimation using ctffind3 (right).



Individual particles (odd rows) and re-projections of the model at the assigned orientations (even rows) produced by FREALIGN.

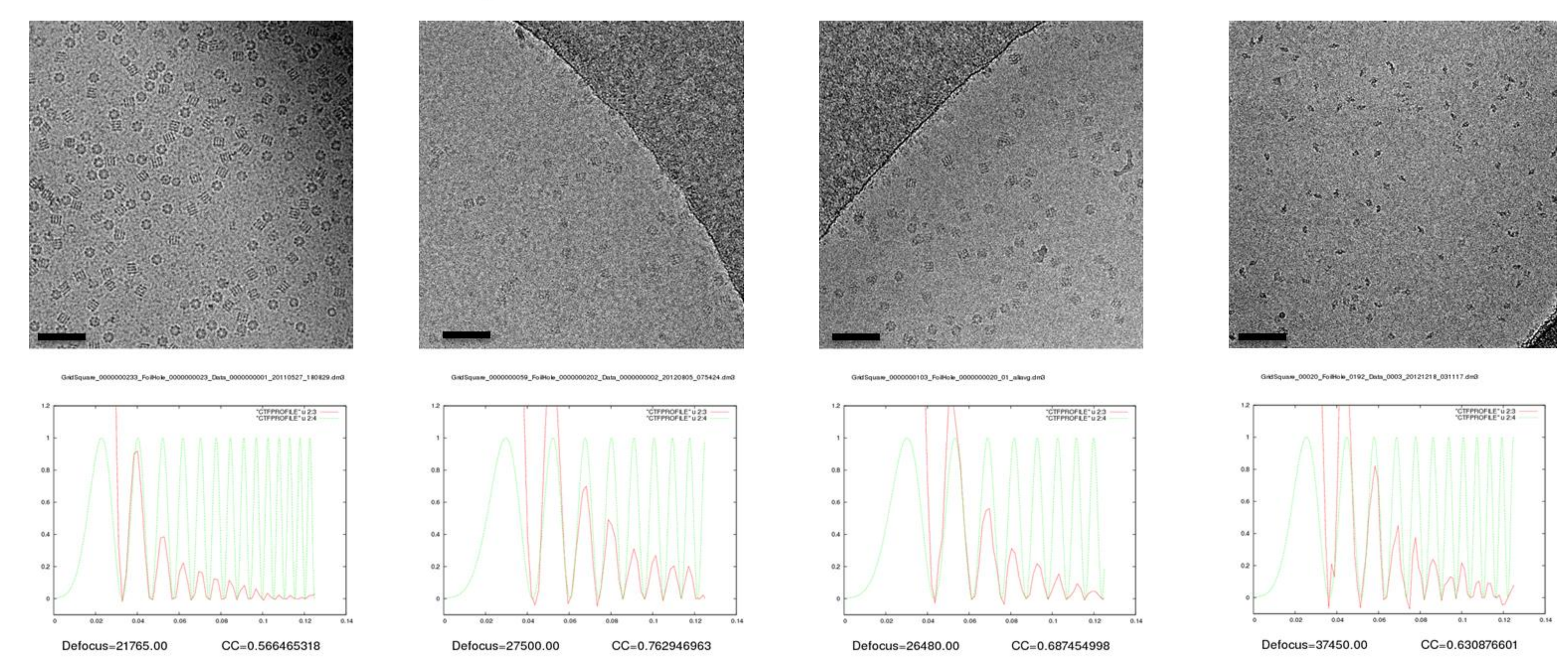
Assessing Multiple Imaging Schemes in Cryo-EM

- **GroEL and β-galactosidase Imaged at Different Accelerating Voltages, Using Different Detectors and Dose Fractionation Schemes**

	Sample	kV	Detector	Exposure	Number of Micrographs	Particles
1	GroEL	80	CCD	1 @ 10 e⁻/Å²	2296	101765
2	GroEL	300	Falcon	1 @ 20 e⁻/Å²	3576	177264
3	GroEL	300	Falcon	3 @ 7 e⁻/Å²	971	34631
4	BGal	300	Falcon	1 @ 20 e⁻/Å²	1946	126387

- All datasets acquired using an FEI Titan Krios.

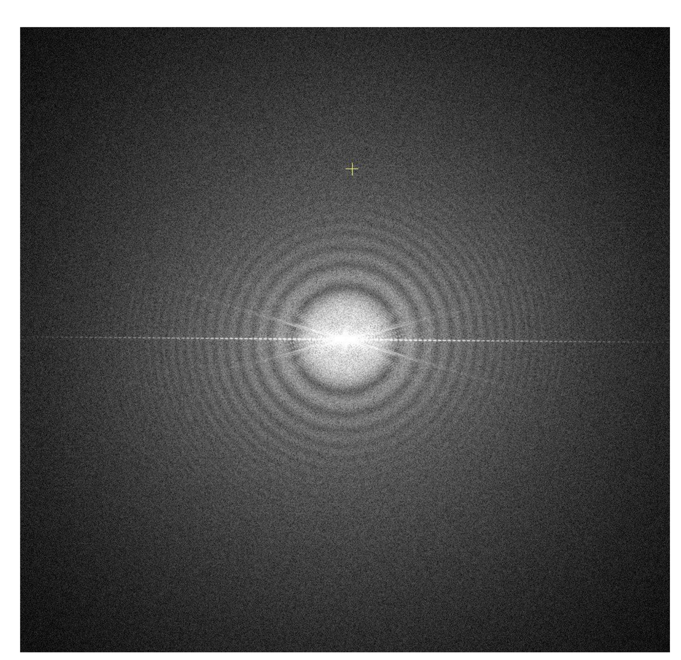
- **Representative Images From Each Dataset (1-4)**



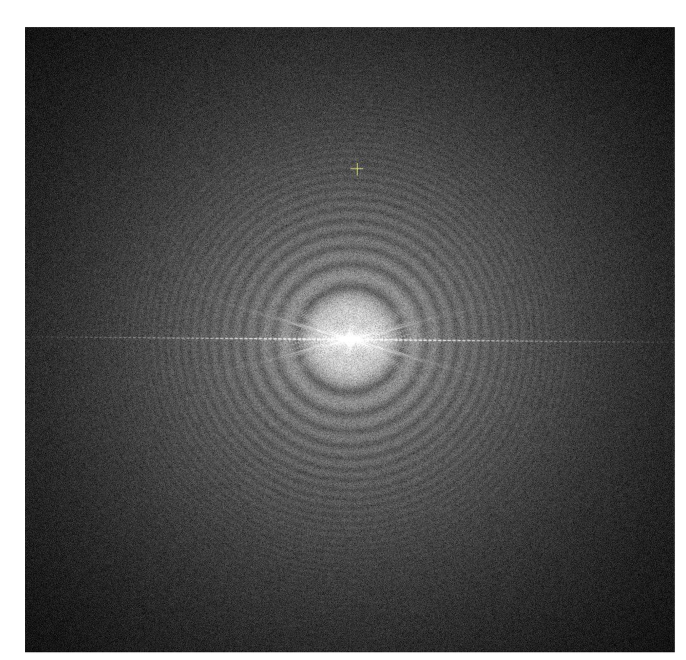
Micrographs and corresponding radially averaged Power Spectra as a function of resolution. Scale bars 6nm.

Effects of Motion Correction

- **Recording of successive exposures combined with image alignment significantly improves image quality**

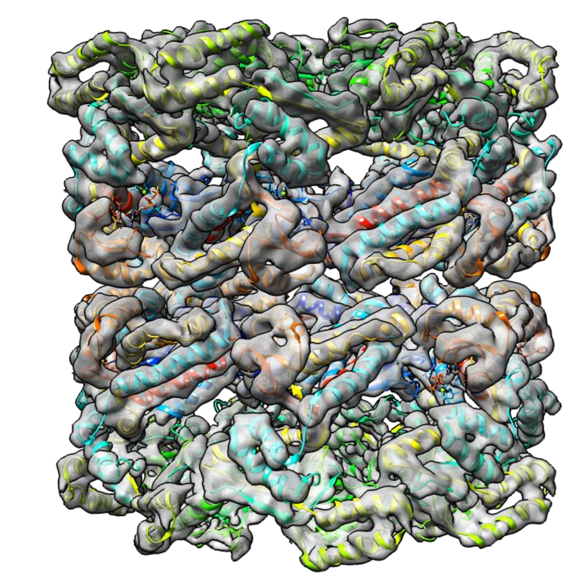


Power Spectrum of average of three successive exposures without accounting for motion effects.

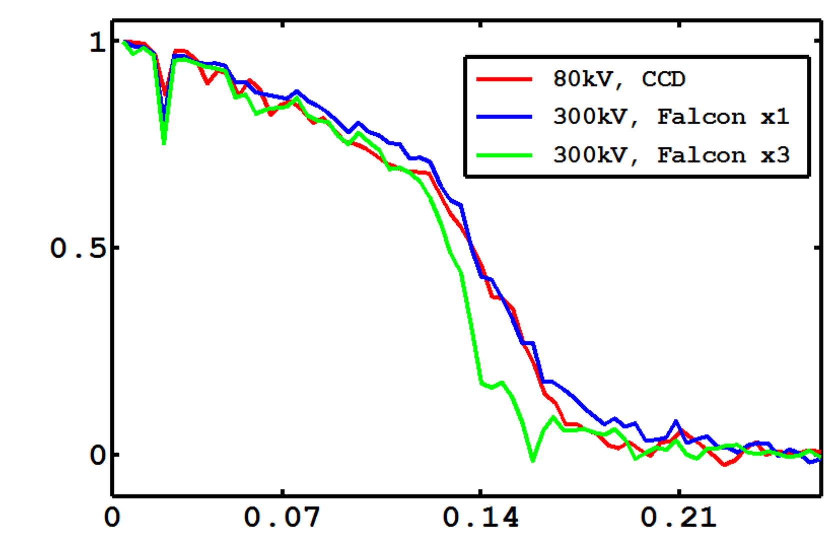


Power Spectrum of average after motion correction showing better resolved Thon rings throughout image.

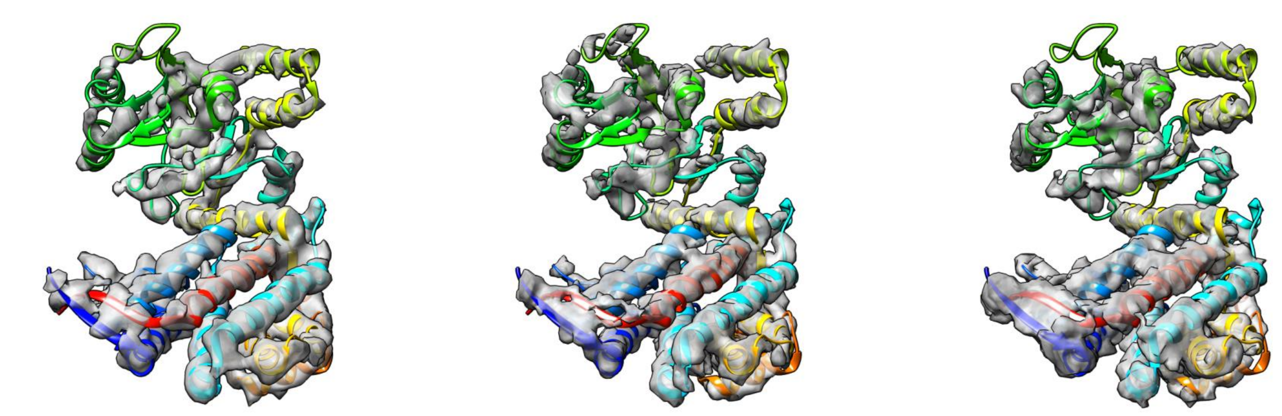
Reconstructions At Sub-Nanometer Resolution



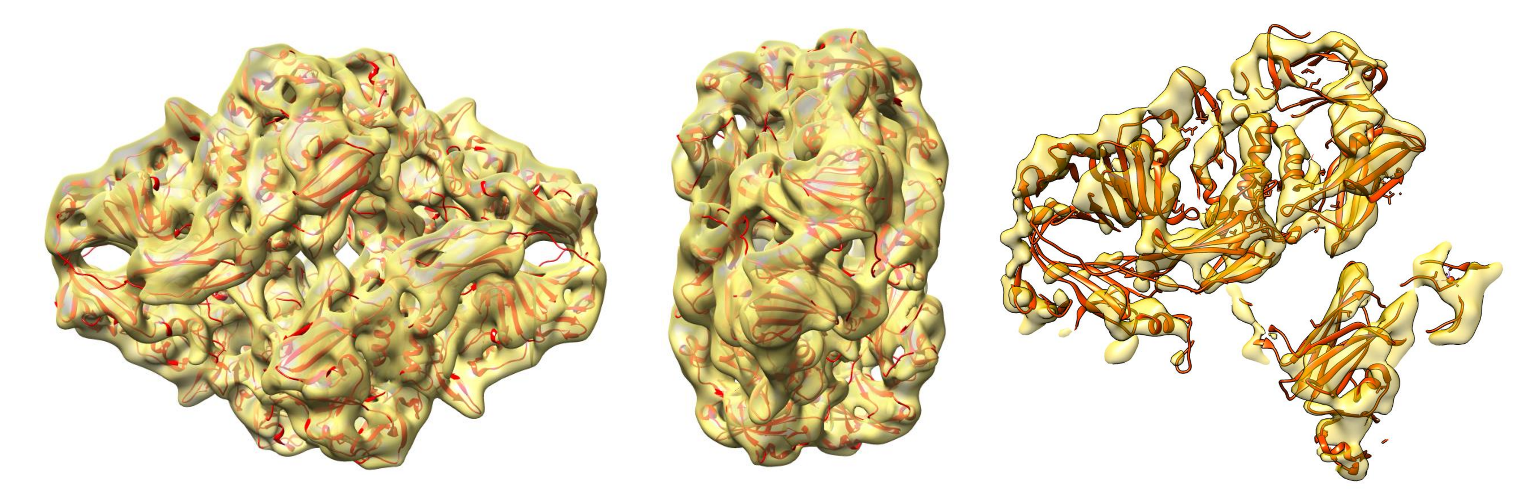
Reconstruction from ~20,000 GroEL particles imaged at 80kV using a CCD detector.



FSC plots against X-ray model (PDB ID. 3e76) shows ~7Å resolution (.5 cutoff) obtained from datasets 1-3.



Asymmetric sub-unit of GroEL obtained from datasets 1-3 with fitted X-ray coordinates (PDB ID. 3e76).



Reconstruction of β-galactosidase obtained from ~30,000 particles using an initial model obtained from sub-tomogram averaging. Fitted X-ray coordinates (PDB ID. 3VD3).

Summary

We present automated workflows for data collection and processing of large single-particle datasets suitable for high-throughput data analysis. We applied these workflows to protein samples imaged under a variety of cryo-EM conditions showing that reconstructions at sub-nanometer resolution can be consistently achieved using our strategy.

Acknowledgments

This study was performed within "The Living Lab for Structural Biology", a collaboration between NIH and FEI (http://livinglab.nih.gov/) and utilized the high-performance computational capabilities of the Biowulf Linux cluster at the National Institutes of Health, Bethesda, MD (http://biowulf.nih.gov/).